

Disfluences en parole continue en français : paramètres prosodiques des répétitions

Ivana Didirková¹ Yaru Wu² Anne-Catherine Simon³

(1) UR 1569 TransCrit, Université Paris 8 Vincennes - Saint-Denis, France

(2) UR 4255 CRISCO, Université de Caen Normandie, France

(3) Centre VALIBEL, Institut Langage et Communication,
Université catholique de Louvain, Belgique

ivana.didirkova@univ-paris8.fr, yaru.wu@unicaen.fr,
anne-catherine.simon@uclouvain.be

RÉSUMÉ

Cette étude s'intéresse aux caractéristiques acoustiques des répétitions à l'identique à travers différents genres de parole en français. Le corpus d'étude LOCAS-F inclut 42 échantillons de parole représentant 14 activités de parole (ou genres). Ces échantillons sont caractérisés en fonction du degré de préparation, d'interactivité, etc. Le nombre de fois que les éléments sont répétés ne varie pas sensiblement en fonction du degré de préparation ou d'interactivité du discours. Du point de vue des durées, les résultats montrent que la durée totale des répétitions augmente avec le degré de préparation du discours. Cela vaut aussi pour la durée des séquences de répétitions avec les insertions. Du point de vue de la fréquence fondamentale, les écarts mélodiques au début d'une séquence de répétition sont plus étendus dans la parole spontanée (non préparée).

ABSTRACT

Disfluencies in continuous speech in French : prosodic parameters of repetitions

This study focuses on the acoustic characteristics of identical repetitions across different speech genres in French. The LOCAS-F study corpus includes 42 speech samples representing 14 speech activities (or genres). These samples are characterised according to the degree of preparation, interactivity, etc. The number of times elements are repeated does not vary significantly according to the degree of preparation or interactivity of the speech. From the point of view of duration, the results show that the total duration of repetitions increases with the degree of preparation of the speech. This is also true for the duration of sequences of repetitions with insertions. In terms of fundamental frequency, the melodic gaps at the start of a repetition sequence are wider in spontaneous (unprepared) speech.

MOTS-CLÉS : répétitions, disfluences, genres de parole, durée, fréquence fondamentale.

KEYWORDS: repetitions, disfluencies, speech genres, duration, pitch.

1 Introduction

Cette contribution traite des caractéristiques acoustiques des répétitions dans un corpus multigenre en français parlé. Elle vise à mieux décrire les phénomènes de répétitions en comparant différentes situations de parole. L'objectif est donc triple : décrire la forme des répétitions présentes dans le corpus dans son ensemble, analyser leur distribution en fonction des paramètres situationnels (comme le degré de préparation) et enfin mesurer les caractéristiques acoustiques des répétitions, en

particulier leur durée et leur fréquence fondamentale. Nous nous intéresserons particulièrement aux traits distinguant la parole préparée/non préparée et interactive/non interactive. En effet, nous nous attendons à ce que ces paramètres influencent la fréquence ou la forme des répétitions : les répétitions devraient être moins fréquentes dans un discours préparé (moins d'effort de planification) et plus courtes dans un discours peu interactif (moins de risque d'interruption). À plus long terme, ce type d'études permet de décrire quels types de répétitions sont "normales" (norme objective) en fonction de différentes situations de parole.

L'introduction permet de définir différents types de répétitions (1.1), d'explicitier la terminologie retenue pour l'annotation des données (1.2) et de présenter les principales caractéristiques acoustiques des répétitions relevées dans la littérature existante (1.3).

1.1 Différents types de répétitions

Traditionnellement, les répétitions sont considérées comme des disfluences ou des marques d'hésitation. Cependant, des travaux ont montré que toute répétition qu'on observe dans la langue parlée n'est pas disfluente. [Blanche-Benveniste et al. \(1990, p. 18-22\)](#) sont parmi les premiers à décrire de manière systématique les différentes formes de piétinement qui peuvent jalonner la progression de la parole. « Le déroulement syntagmatique est brisé [...] lorsque le locuteur répète [...]. Aucune règle du français ne pourrait nous inciter à enchaîner deux occurrences de *hier* pour y trouver une relation de dépendance » (p. 18). La répétition est un piétinement sur une seule et même place syntaxique. Si la plupart des répétitions sont disfluentes, certaines visent cependant un effet stylistique. Dans les exemples (1) à (3), les répétitions de *nous* et de *non* illustrent respectivement (1) une répétition à fonction syntaxique, chaque élément occupant une fonction différente dans la dépendance du verbe ; (2) une répétition disfluente, actualisant une hésitation ; (3) une répétition qu'on peut potentiellement qualifier d'intensive, et donc d'intentionnelle.

1. une fois par année **nous nous** réunissons et nous célébrons notre patrie sans fausse modestie (corpus LOCAS-F, pol-3)
2. **nous nous** vou/ **nous** vivons une époque catastrophique dans tous les domaines (corpus LOCAS-F, intlib-3)
3. **non non** il avait l'air hyper clean le gars (corpus LOCAS-F, conv-i-3)

Si les répétitions à fonction syntaxique (1) doivent être écartées, il n'est pas toujours aisé de distinguer entre les répétitions disfluentes et les répétitions stylistiques ou expressives ([Blanche-Benveniste et al., 1990, p. 21](#)). Par conséquent, la plupart des études annotent l'ensemble de ces répétitions dans une seule catégorie ([Crible et al., 2015](#)).

1.2 Terminologie

Une répétition est une disfluence complexe, composée de plusieurs éléments. Le modèle de [Shriberg \(1999\)](#) permet de décrire de manière générale la structure de la plupart des disfluences qui s'observent dans la parole. Le premier élément de la zone disfluente est le *reparandum*, le matériel verbal qui sera ultérieurement remplacé. La fin de cet élément marque l'interruption de la fluence verbale, dans la mesure où le locuteur a détecté un problème qui empêche le déroulement fluide de la production en cours. La phase d'édition se trouve entre l'élément à corriger (*reparandum*) et la réparation elle-même (*repair*), et elle peut contenir des éléments comme une pause silencieuse, une pause pleine ou des termes de type *enfin, je veux dire*, etc. La phase d'édition peut également être vide. Enfin, la réparation signale la reprise d'une parole fluente et la continuation du processus de production de la parole.

Ce modèle permet de distinguer, au sein d'une séquence formant une répétition, l'élément *répétable* et un deuxième élément, identique au premier, qui est nommé *répété* (Candea, 2000, p. 315). L'élément répétable peut être répété deux ou trois fois et ces différents éléments peuvent être séparés par d'autres marques du travail de formulation, comme des pauses silencieuses ou des pauses pleines de type *euh*.

1.3 Caractéristiques des hésitations

Comme d'autres auteurs, Grosjean & Deschamps (1975) constatent que les répétitions ne sont pas les marques d'hésitation les plus fréquentes parmi les « variables secondaires » de la parole, à savoir les pauses remplies, les syllabes allongées, les répétitions et les faux départs. Les répétitions représentent 16,80% de ces marques secondaires et deux tiers d'entre elles (64,22%) touchent des mots grammaticaux. En outre, 24,77% des répétitions concernent plusieurs mots. Le fait que les répétitions touchent davantage les mots grammaticaux que les mots lexicaux est rapporté par la plupart des études empiriques. Candea (2000, p. 316) observe des répartitions similaires à celles rapportées par Grosjean & Deschamps (1975) : les répétitions concernent 2,94% des mots outils et 1,43% des ligateurs (élément qui précise le lien avec ce qui précède, selon Morel & Danon-Boileau 1998, p. 20, cités par Candea 2000, p. 68) font l'objet d'une répétition, pour seulement 0,56% des mots lexicaux. Dans le corpus de Grosjean & Deschamps (1975, p. 180), on observe que 49% des répétitions sont précédées par une pause (silencieuse, pleine ou allongement). Candea (2000, p. 325) montre que la plupart des répétitions se combinent avec un allongement de l'élément répétable (53,93%). En outre, dans 82,42% des cas, la durée du mot répété est supérieure à celle du mot répétable (p. 335). Lorsqu'elles sont simples, sans combinaison avec une autre marque d'hésitation, les répétitions de mots outils sont très brèves (max. 90 s). Cela incite l'auteure à les considérer comme une forme de retard articulatoire de la production vocale, une forme de léger bégaiement (p. 327-328). Lorsqu'elles se combinent avec d'autres marques du travail de formulation, comme une pause silencieuse, un allongement ou une pause pleine de type *euh*, les répétitions forment des séquences plus longues qui peuvent approcher les 3 secondes. La combinaison avec un *euh* représente 39,6% des cas. Grosman (2018, p. 252) constate que les répétitions à l'identique forment la troisième marque la plus fréquente de disfluente, après les pauses silencieuses et les pauses pleines. Elles se produisent à une fréquence de 18,9 répétitions pour 1000 mots et représentent 5,5% du temps d'articulation. Analysant un corpus multigenre, elle observe que le degré de préparation est une caractéristique qui prédit un plus faible taux de répétitions (p. 254). Près de la moitié (45,83%) des répétitions sont produites de manière simple, sans autre terme d'édition. Lorsqu'une combinaison se produit, c'est avec une pause silencieuse qu'elle est la plus fréquente (29,7%), puis avec un marqueur de discours (10,1%) et enfin avec une troncation (9,6%) (Grosman, 2018, p. 259). Les données de Grosman ne permettent pas de confirmer le fait que l'élément répétable est généralement plus long que l'élément répété, c'est le cas dans seulement 11% des cas (p. 270).

2 Méthode

2.1 Corpus et annotation

Le corpus LOCAS-F contient 42 échantillons de parole représentant 14 activités de parole pour une durée totale de 3h59 minutes et un nombre de 41 322 tokens. Les répétitions ont été annotées manuellement selon un protocole mis au point par Crible *et al.* (2015). Dans cette étude, les échantillons contenant des narrations conversationnelles ont été écartés en raison des nombreux chevauchements qui auraient pu fausser les mesures de f_0 . Nos analyses portent donc sur 2h38min35sec d'enregistrements (durée moyenne des enregistrements : 4min53). Les répétitions font partie des marques de

disfluences composées, comportant plus d'un élément. Tous les types de répétitions (cf. 1.1), disfluents ou stylistiques, sont annotés. Les répétitions à l'identique (RI) couvrent « un mot ou une séquence de mots (quasi-)contigus répétés formellement à l'identique » (p. 14). La quasi-contiguïté rend compte de la possibilité d'insérer un ou plusieurs éléments disfluents entre l'élément répétable et l'élément répété, comme une pause vide (UP), une pause pleine de type *euh* (FP), un marqueur de discours (DM), un terme d'édition (ET) voire une insertion parenthétique (IP). Les éléments répétés sont numérotés de 0 (élément répétable) à N (en fonction du nombre de répétés). Les répétitions avec modulation (RM) font varier un élément et les répétitions grammaticales impliquent une forme différente appartenant à la même catégorie grammaticale.

La fiabilité de l'annotation a été vérifiée au moyen d'un score d'accord interannotateurs portant sur 25,7 minutes de parole et incluant des parties du corpus LOCAS-F. « L'accord interannotateurs obtenu sur l'ensemble des données s'élève à $K = 0,67$ (ratés = 2 ; $z = 98,60$; $p < 0,001$; $n = 6892$), ce qui correspond à un accord fort » (Grosman, 2018, p. 120).

Chaque échantillon de parole est également annoté en fonction de critères situationnels : le degré de préparation, le degré d'interactivité, le caractère public ou privé, médiatique ou non médiatique, etc. Concernant le degré de préparation, qui est un paramètre important pour notre étude, l'annotation comprend trois degrés : un discours est non préparé ou spontané lorsque le locuteur l'improvise au fur et à mesure et n'en prépare aucune partie par écrit ; un discours est semi-préparé lorsque le locuteur prépare le contenu global du discours sans le rédiger in extenso (la production orale peut inclure un support écrit partiel) ; le discours est préparé lorsqu'il a été entièrement rédigé par écrit, qui soit lu ou récité lors de la production orale.

2.2 Traitement des données

Différents types de répétitions sont disponibles dans nos données. Il s'agit plus spécifiquement de répétitions dites à l'identique (RI), où tous les éléments répétés sont les mêmes (par ex. « j'arrive dans j'arrive dans 5 minutes »), qu'elles soient ou non interrompues par un / plusieurs autres éléments (« j'arrive dans euh j'arrive dans 5 minutes »), et de répétitions dites avec modulation (RM), où la structure des éléments répétés est préservée mais au moins un élément sera modifié (« j'ai vu ton j'ai vu ta cousine »).

Dans cette étude, nous avons fait le choix de nous concentrer sur les séquences de mots répétés à l'identique (RI) et ce, afin d'éviter l'influence de la modification sur la longueur de l'élément répété et son comportement en termes de f_0 .

La première partie des analyses porte sur les répétitions à l'identique, avec ou sans éléments insérés, et vise notamment à décrire ces dernières en termes de durée (en tenant compte ou non des insertions) et en termes de nombre d'éléments répétés. Ces mesures sont d'abord analysées en fonction du degré de préparation du discours, puis en fonction du degré d'interactivité. En effet, nous supposons qu'un discours préparé devrait être moins sujet aux répétitions que des discours semi- ou non-préparés, puisque nécessitant moins de planification de la part du locuteur. D'un autre côté, les discours avec un degré d'interactivité plus important devraient favoriser des répétitions longues, afin de signaler à l'interlocuteur la volonté de continuer le discours et éviter ainsi une interruption par l'interlocuteur.

La seconde partie de nos analyses porte sur la fréquence fondamentale des répétitions. Pour mieux se concentrer sur la comparaison de la fréquence fondamentale entre les séquences de répétition et les syllabes environnantes, nous avons décidé de nous concentrer sur les séquences de répétition avec des mots qui ne sont répétés qu'une seule fois. Plus précisément, nous nous intéressons aux

intervalles mélodiques entre la première syllabe du mot avant sa répétition et la dernière syllabe du mot qui le précède (« RI0 - PrecSyl ») et aux intervalles mélodiques entre la première syllabe du mot suivant et la dernière syllabe du mot répété pour la première fois (« FollSyl - RI1 »), par rapport aux intervalles mélodiques entre les autres syllabes (« Autres »). Par exemple, la séquence de mots « bah la question » a été produite avec le mot « la » répété une seule fois (« *bah la_(R0) la_(RI) question* »). Nous nous intéressons donc aux intervalles mélodiques entre la première syllabe du mot avant sa répétition « la_(R0) » ([la]) et la dernière syllabe de « bah » ([ba]) et aux intervalles mélodiques entre la première syllabe du mot « question » ([kɛs]) et la dernière syllabe du mot répété pour la première fois « la_(RI) » ([la]). Nous avons analysé environ 29k tokens pour chacune des analyses d’intervalles mélodiques (dont 225 « RI0 - PrecSyl » et 216 « FollSyl - RI1 »).

Les valeurs de fréquence fondamentale ont été extraites en utilisant Praat (Boersma & Weenink, 2024) et les valeurs mesurées en Hz ont été converties en demi-tons (DT) par rapport à 50 Hz.

Les syllabes ont été exclues de toute analyse lorsqu’elles sont indéfinies en termes de mesure de fréquence fondamentale (« undefined ») ou lorsque la syllabe qui les précède ou les suit est une pause remplie (« UP »).

3 Analyses et résultats

3.1 Nombre de répétitions

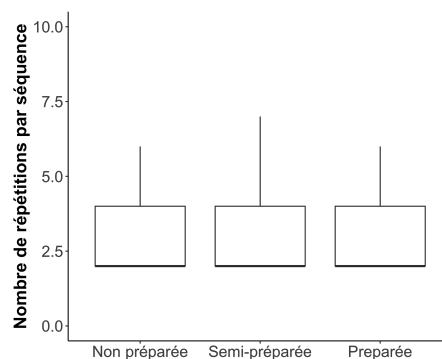


FIGURE 1 – Nombre de répétitions par séquence

La figure 1 montre qu’il n’y a pas d’effet du degré de préparation sur le nombre total d’éléments répétés dans chaque séquence analysée. En effet, dans les discours préparés, en moyenne, les locuteurs répètent 3,2 fois chaque élément, alors qu’ils les répètent en moyenne 2,92 fois dans les discours semi-préparés et 3,04 fois dans les discours non-préparés. Le constat est le même avec le degré d’interaction où aucun effet de ce dernier n’est observé sur la même variable, avec des éléments répétés 3,25 fois en moyenne dans les discours interactifs, 2,92 fois dans les discours semi-interactifs et enfin 2,91 fois en moyenne dans les discours non-interactifs.

3.2 Durée

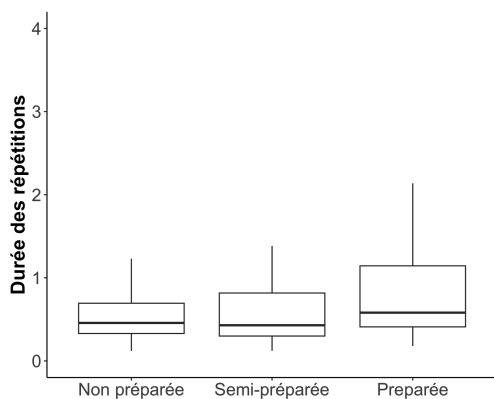


FIGURE 2 – Durée des répétitions en fonction du degré de préparation.

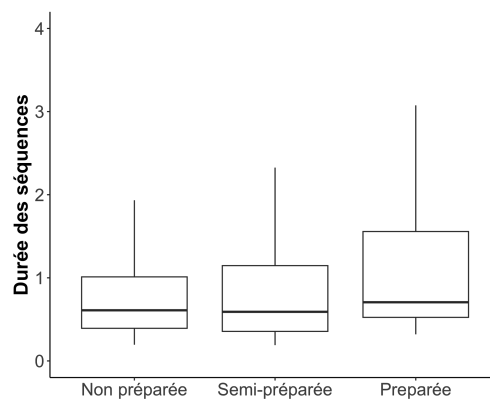


FIGURE 3 – Durée des séquences de répétition en fonction du degré de préparation.

Lorsque l'on s'intéresse à la durée totale des répétitions (cf. Figure 2, hors insertions éventuelles, c'est-à-dire uniquement la durée des éléments répétés à l'identique), l'on s'aperçoit que cette durée varie de manière significative en fonction du degré de préparation. Plus concrètement, les répétitions sont plus longues dans les discours préparés (0,92s en moyenne, ET = 0,75s, $p = 0,004$, $t = 2,981$, SE = 0,112) que dans les discours non-préparés (moyenne = 0,58s, ET = 0,43s). Il n'y a pas de différence significative entre les discours semi-préparés (moyenne : 0,64s, ET = 0,53s) et non-préparés. En revanche, ce même paramètre, à savoir la durée totale des répétitions, ne semble pas être influencé par le caractère interactif ou non des discours (non-interactif : durée moyenne 0,72s, ET = 0,58s ; semi-interactif : durée moyenne 0,58s, ET = 0,47s ; interactif : durée moyenne 0,59s, ET = 0,43s).

Enfin, comme le montrent les LMM, il existe également une variation significative de la durée totale de la séquence (répétitions + éventuelles insertions) en fonction du degré de préparation d'un discours (Figure 3). En effet, la séquence répétée est significativement plus longue dans les discours préparés (moyenne = 1,22s, ET = 1s) que dans les discours non-préparés (moyenne = 0,81s, ET = 0,64s, $p = 0,03$, $t = 2,233$, SE = 0,174). Il n'y a pas de différence significative avec les discours semi-préparés (moyenne = 1s, ET = 0,98s). Si l'on s'intéresse à la durée totale de la séquence répétée en fonction du degré d'interactivité, l'on s'aperçoit que les discours non-interactifs entraînent une durée de la séquence disfluente plus longue (moyenne = 1,13s, ET = 1,06s) que les discours interactifs (moyenne = 0,83s, ET = 0,66s, $p = 0,04$, $t = 2,107$, SE = 0,125). Ici encore, il n'y a pas de différence significative avec les discours semi-interactifs (moyenne = 0,83s, ET = 0,69s).

3.3 Intervalle mélodique

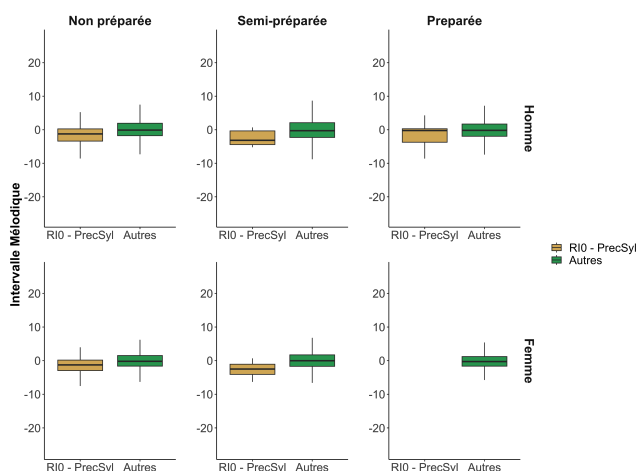


FIGURE 4 – Intervalle mélodique (en DT) entre la dernière syllabe du mot précédent et la première syllabe du mot en RI (RI0), en fonction du degré de préparation et du genre (de gauche à droite : parole non préparée, semi-préparée et préparée ; partie en haut pour les hommes, partie en bas pour les femmes).

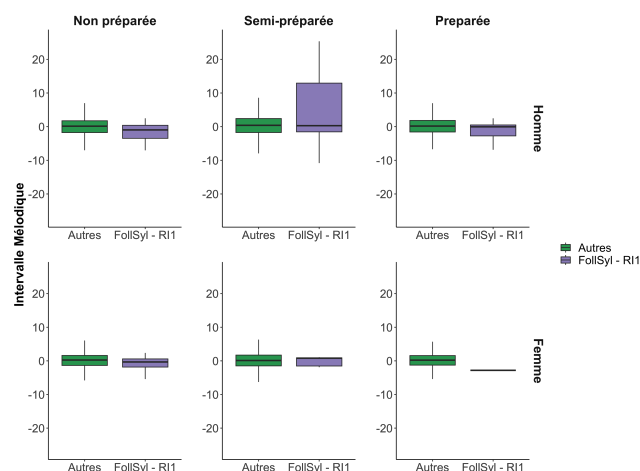


FIGURE 5 – Intervalle mélodique (en DT) entre la dernière syllabe du mot répété (RI1) et la première syllabe du mot suivant, en fonction du degré de préparation et du genre (de gauche à droite : parole non préparée, semi-préparée et préparée ; partie en haut pour les hommes, partie en bas pour les femmes).

Les résultats sur les intervalles mélodiques sont présentés dans la figure 4 et la figure 5, pour les hommes (en haut) et les femmes (en bas) et pour les trois degrés de préparation (de gauche à droite : parole non préparée, semi-préparée et préparée).

Comme nous pouvons le voir dans la figure 4, il n'y a pas de données de locutrice en parole préparée dans l'ensemble de données sélectionné. L'intervalle mélodique est plus faible pour « RI0-PrecSyl » que pour « Autres » en général, pour les locuteurs masculins et féminins, ainsi que pour les trois différents degrés de préparation. Les résultats basés sur le modèle LMM confirment que l'intervalle mélodique est significativement plus faible pour « RI0-PrecSyl » ($\beta = -5,671e-01$; $t = 2,120$; $SE = 2,675e-01$) par rapport à celui observé pour « Autres ». L'intervalle mélodique est significativement plus élevé pour la parole non-préparée ($\beta = 3,076e-01$; $t = 3,755$; $SE = 8,193e-02$), comparé à celui observé pour la parole préparée. Aucune différence significative n'est observée entre les paroles préparée et semi-préparée, ni entre les hommes et les femmes.

Des tendances similaires sont observées pour les hommes et les femmes pour la parole non préparée (première colonne à partir de la gauche) et pour la parole préparée (troisième colonne à partir de la gauche) dans la figure 5 : le changement d'intervalle mélodique est plus faible pour « FollSyl-RI1 » que pour « Autres ». Pour la parole semi-préparée, aucune différence significative n'est observée. Les résultats basés sur le modèle LMM confirment que l'intervalle mélodique est significativement plus faible pour « FollSyl-RI1 » ($\beta = -1,470e+00$; $t = 5,107$; $SE = 2,879e-01$) par rapport à celui observé pour « Autres ». L'intervalle mélodique est significativement plus élevé pour la parole non préparée ($\beta = -1,734e-01$; $t = 2,620$; $SE = 6,620e-02$), comparé à celui observé pour la parole préparée. La différence n'est pas significative entre les paroles préparée et semi-préparée, ni entre les hommes et les femmes.

4 Discussion, limites et perspectives

L'objectif de cette étude était de mieux décrire les répétitions identiques en parole continue en français du point de vue de leur durée, du nombre d'éléments répétés et de la fréquence fondamentale, le tout en corrélation avec le degré de préparation du discours (préparé, semi-préparé, non-préparé) et le degré d'interactivité (interactif, semi-interactif, non-interactif).

Les résultats portant sur le nombre d'éléments répétés ne montrent aucune corrélation significative entre cette variable et le degré de préparation ou d'interactivité d'un discours.

En ce qui concerne la durée totale des répétitions, celle-ci est plus importante lors d'un discours préparé que lorsque le discours présente davantage de spontanéité. Ce résultat peut paraître surprenant : l'on pourrait s'attendre à ce qu'un discours préparé ne nécessite pas de séquences disfluentes particulièrement longues, a fortiori si l'on considère qu'une disfluence résulte d'un manque de préparation de la séquence à suivre. Cependant, [Simon \(2021\)](#) a montré que les répétitions à l'identique sont une manière fréquente d'hésiter dans des discours extrêmement solennels. De même, on pourrait expliquer cette durée quasiment doublée de la répétition dans les discours préparés par le fait que le locuteur, seul, n'a pas besoin d'accélérer sa production de la parole puisqu'il n'est pas concurrencé par un interlocuteur qui pourrait l'interrompre ([Hirsch et al., 2016](#)). Enfin, ces durées accrues des répétitions et des séquences de répétitions dans les discours préparés et non-interactifs pourraient s'expliquer par une corrélation avec la variation du débit, plus lent dans les genres plus formels ([Simon et al., 2010](#)).

Enfin, l'étude sur l'intervalle mélodique montre que la différence entre la syllabe précédant le premier élément répété et la première syllabe du mot répété est plus faible qu'entre des syllabes qui ne sont pas concernées par les répétitions. La même tendance est observée entre la dernière syllabe de la répétition et la première syllabe du mot suivant. Ce résultat montre que les premières et dernières syllabes des éléments répétés seraient davantage intégrées prosodiquement, autorisant moins de sauts de f_0 que d'autres éléments présents dans les discours. Si une tendance similaire a été observée dans notre précédente étude sur les allongements et les pauses pleines ([Wu et al., 2022](#)) entre la disfluence et la syllabe suivante, le résultat montrant que la première syllabe de la répétition est plus intégrée prosodiquement que la moyenne indique un comportement différent des répétitions comparé aux deux autres types de disfluences. Ce constat serait à approfondir en analysant les intervalles mélodiques entre les différents éléments répétés (par exemple entre deux « je suis je suis »).

Nos analyses présentent un certain nombre de limites. Tout d'abord, nous avons, pour cette étude, exclu les répétitions comprenant des modifications d'un ou plusieurs éléments. Or, il serait intéressant de comparer leur fréquence fondamentale à celle des types de disfluences semblables, à savoir les reprises et les répétitions à l'identique. Cette comparaison permettrait de savoir si les répétitions modifiées se rapprochent davantage de l'un ou de l'autre type, ou si elles se situent à mi-chemin entre les deux. En effet, les répétitions avec modification (auto-correction) montrent une capacité de contrôle de la production orale, dont le degré varie probablement selon la situation de parole. Nous avons également exclu les cas complexes de répétitions à l'identique. Il s'agit notamment de cas où plusieurs répétitions d'un élément se retrouvent entrecoupés par des insertions et sont modifiées avant d'être reprises à l'identique ("je je euh je euh je suis euh j'étais je suis") et pour lesquelles une catégorie spécifique devrait être créée.

Enfin, nous entendons poursuivre nos analyses notamment en incluant d'autres types de disfluences (faux-départs, reprises, etc.) et en comparant les paramètres étudiés dans ces différents types de disfluences, dans l'objectif de déterminer s'il existe des différences saillantes en fonction du type de la disfluence (voir aussi l'étude de [Wu et al., 2022](#)).

Références

- BLANCHE-BENVENISTE C., MERTENS P. & WILLEMS D. (1990). Le français parlé : études grammaticales. *Éditions du CNRS*.
- BOERSMA P. & WEENINK D. (2024). Praat : doing phonetics by computer [computer program]. <http://www.praat.org/>.
- CANDEA M. (2000). *Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané. Etude sur un corpus de récits en classe de français*. Thèse de doctorat, Université de la Sorbonne nouvelle-Paris III.
- CRIBLE L., DUMONT A., GROSMAN I. & NOTARRIGO I. (2015). " annotation des marqueurs de fluence et disfluence dans des corpus multilingues et multimodaux, natifs et non natifs v. 1.0.
- GROSJEAN F. & DESCHAMPS A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français : vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, **31**(3-4), 144–184.
- GROSMAN I. (2018). *Évaluation contextuelle de la (dis) fluence en production et perception : pratiques communicatives et formes prosodico-syntaxiques en français*. Thèse de doctorat, UCL-Université Catholique de Louvain.
- HIRSCH F., MARSAC F., DIDIRKOVA I., BECHET M. & MESSAOUD M. A. B. (2016). Spécificités du rythme de la parole politique. le cas de François Hollande. *Romanica Wratislaviensia*, p. p–145.
- MOREL M.-A. & DANON-BOILEAU L. (1998). *Grammaire de l'intonation l'exemple du français*. Editions Ophrys.
- SHRIBERG E. (1999). Phonetic consequences of speech disfluency. In *Proceedings of the International Congress of Phonetic Sciences*, volume 1, p.2.
- SIMON A.-C. (2021). De l'écrit à l'oral dans les discours solennels : le cas de l'académie française. *Linguistique de l'écrit*, **2**, éparants.
- SIMON A.-C., AUCHLIN A., AVANZI M., GOLDMAN J.-P. *et al.* (2010). Les phonostyles : une description prosodique des styles de parole en français. *Les voix des Français : en parlant, en écrivant*, Bern : Lang, p. 71–88.
- WU Y., DIDIRKOVÁ I. & SIMON A.-C. (2022). Disfluences en parole continue en français : paramètres prosodiques des pauses pleines et des allongements vocaliques. In *Proc. XXXIVe Journées d'Études sur la Parole – JEP 2022*, p. 900–909. DOI : [10.21437/JEP.2022-95](https://doi.org/10.21437/JEP.2022-95).